Are non-protein coding RNAs junk or treasure?
An attempt to explain and reconcile opposing viewpoints of whether the human genome is mostly transcribed into non-functional or functional RNAs

Nils G. Walter

Center for RNA Biomedicine, Single Molecule Analysis Group, Department of Chemistry, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
Nils G. Walter, Center for RNA Biomedicine, Single Molecule Analysis Group, Department of Chemistry, University of Michigan, Ann Arbor, MI 48109-1055, USA.
Email: nwalter@umich.edu

Funding information
NIH, Grant/Award Number: GM131922

Abstract
The human genome project's lasting legacies are the emerging insights into human physiology and disease, and the ascendance of biology as the dominant science of the 21st century. Sequencing revealed that >90% of the human genome is not coding for proteins, as originally thought, but rather is overwhelmingly transcribed into non-protein coding, or non-coding, RNAs (ncRNAs). This discovery initially led to the hypothesis that most genomic DNA is "junk", a term still championed by some geneticists and evolutionary biologists. In contrast, molecular biologists and biochemists studying the vast number of transcripts produced from most of this genome "junk" often surmise that these ncRNAs have biological significance. What gives? This essay contrasts the two opposing, extant viewpoints, aiming to explain their bases, which arise from distinct reference frames of the underlying scientific disciplines. Finally, it aims to reconcile these divergent mindsets in hopes of stimulating synergy between scientific fields.

KEYWORDS
bacterial riboswitches, biological function, evolutionary selection, genetics and genomics, gene expression & regulation, human genome, non-protein coding RNA

INTRODUCTION
As many stories, this essay starts with a personal experience. One fall term, teaching a special topics course on "Nucleic Acid Biochemistry" for senior undergraduates and beginning graduate students, I had asked teams of students to each develop or expand a Wikipedia entry related to our course material. One group identified "Repeated sequence (DNA)" as a rudimentary entry that needed some work. The students added—based on an exhaustive literature search—much flesh to the existing "bones" in the form of the history, types, and functions in physiology and disease of these pervasive repeat sequences. At the end of the term, the students moved the entry from their "Sandbox" to Wikipedia proper.[1] Remarkably, within 24 h an international team of retired and active geneticists and evolutionary biologists had prominently added a sentence to the entry, suggesting that these repeat sequences are likely non-functional and belong to the "junk" or "selfish" DNA of the cell. Additionally, they published an accompanying blog about their grievance that the term "junk DNA" had been removed by the students.[2]

Personally, this strong reaction surprised me as I had assumed that modern transcriptome analyses, funded in part since 2004 by the National Human Genome Research Institute (NHGRI) through the ENCODE project,[3–8] had long dispelled the early misperception that our genome harbors 90% useless "junk DNA", since most of it is transcribed into non-protein coding, or non-coding, RNAs (ncRNAs, both small and long, that is, lncRNAs; Figure1).[9] Through a long exchange
The definition of scientific terms is especially important for debates between different fields since each area of science may use a distinct frame of reference. Case in point—while the term "biological function" refers to one of the most foundational tenets of biology, it turns out to be differently defined by geneticists and evolutionary biologists versus molecular biologists and biochemists. For those biosciences studying whole organisms or ecosystems, a biological function describes the reason why an organism harbors a particular trait or behavior. In contrast, the molecular biologist and biochemist will think of biological function as a specific biochemical activity that a molecule carries out within the cells of the organism. Such a function may be as rudimentary as the recruitment of a protein by a specific segment of a nucleic acid. Accordingly, the ENCODE team defined RNA elements as functional when they bound one of the plethora of RNA-binding proteins (RBPs) found in the mammalian cell. How that protein-binding RNA element fits into, for example, a specific regulatory pathway and how this pathway then leads to a particular trait or behavior of the organism as a whole is of secondary concern.

A related term that critically needs definition is that of a "gene". Generally, a gene is defined as a region of DNA that contains instructions for the function, growth and reproduction of an organism that is genetically inherited by the next generation of the organism. While classic, this definition has become more ambiguous as we have learned more about the complexity of multicellular organisms. First, features of the organism as a whole again are used here as central markers...
for defining what constitutes a gene and what does not, excluding ncRNA genes of as-yet unknown function. In addition, a broad interpretation will include the entire non-protein coding regulatory architecture, such as intronic sequences, that instructs the gene to be expressed under a given set of developmental and stress conditions. In addition, genes have fuzzy boundaries since they can overlap and interweave so a broadly inclusive definition of a gene is as a region of DNA that expresses an RNA that may or may not be translated. Second, both geneticists and evolutionary biologists will take conservation of the DNA sequence as an important criterion for defining a gene as functional. This notion certainly holds for housekeeping genes, but runs counter to the observation that the most rapid evolution between the human and other primate genomes occurs outside of protein-coding regions, attributing evolutionary significance to regulatory sequences and RNAs. In addition, single nucleotide alterations between the human genome and that of, for example, chimpanzees constitute ~1.2% of the human genome, with more extended deletions and insertions covering ~3%. These numbers are not very far off from the estimated typical ~0.6% difference between an individual’s genome and the human reference genome. In fact, the human body is a genomic mosaic in that somatic genome variations accumulate during development in response to both genetic programs, which may differ from tissue to tissue, and environmental stimuli, which are often undetected and irreplicable, helping diversify our phenotypes throughout life. Strikingly, modern single-cell sequencing technologies are revealing that the variation among long-living neurons in a single human results from somatic mutations that accumulate in the brain by different mechanisms and rates during development and ageing, and contribute to neuropsychiatric disorders.

Taken together, these observations point to the idea that the conservation of genes may be overrated. That is, while the biological functions of ncRNAs are rooted in simple biochemical activities such as binding of a specific protein, the non-conserved sequence variations between organisms, tissues and individual cells may hold the key to understanding human evolution and complex traits such as human behavior. Just as one example, an estimated ~2300 human microRNA (miRNA) genes control tissue specific gene expression by binding their Argonaute effector proteins (in humans most notably Ago2) to serve as guides towards a large fraction of human messenger RNAs (mRNAs) in such complex combinatorial fashion that the resulting regulatory network functions are challenging to fully delineate. Despite this challenge to exactly define the function of a specific miRNA, the plethora of overlapping miRNA-mediated regulatory effects results in non-linear, emergent properties that empower the plasticity of life and shape much of a human’s phenotype with respect to development, physiology, and disease. More generally, regulatory sequences such as miRNAs tend to have more plasticity than those encoding proteins, enabling them to mutate and evolve rapidly under positive selection for adaptive radiation, when organisms diversify into a new environmental niche.

**HOW DISTINCT FRAMEWORKS ENGENDER DIVERGENT VIEWS OF ncRNAs**

As a biophysical chemist who uses single molecule microscopy, in my respectful, if sometimes antagonistic, electronic discussions with the group of junk DNA champions I naturally adopted the viewpoint that function is a feature of the individual molecule. That is true even though of course only the higher-order interactions between the plethora of diverse molecules inside a cell beget the emergent properties of life. To try to mitigate my own bias, I will refer to the function of an individual molecule as “elemental”, in contrast to the “phenotypical” function that is the higher bar used by geneticists and evolutionary biologists for the definition of the function of, for example, a gene. Based on this difference in definition, critics of ENCODE have pointed out that merely showing the existence of a process, such as transcription of an RNA, or an elemental (or rudimentary, minimal) biochemical activity, such as the transcribed RNA binding a protein, is not sufficient to prove phenotypic (or, in their view, functional) significance. Interestingly, similar controversies exist also at the level of elemental function, for example, where molecular biologists and biochemists who study RNA G-quadruplexes have long debated whether or not these RNA structures exist in vitro, also exist in the human cell and have a discernable function there. Only recently was a specific mechanism discovered by which RNA G-quadruplexes, often found in long ncRNAs, or IncRNAs, downregulate the activity of the polycympressive complex 2 (PRC2) to silence chromatin by installing epigenetic marks on histones. This elemental function of G-quadruplexes maintains gene loci in a transcriptionally active state. Clearly, in this case the absence of evidence did not provide evidence of the absence of function.

On some level, then, the ENCODE controversy is a philosophical disagreement on how much of an ncRNA’s function has to be discovered before legitimately calling it functional, and not junk. The debate was not helped by using the derogatory term “junk”. Yet, embracing the term has been noted that, as more functions of the majority of the human genome are being discovered, ncRNAs have started to move out of the “junkyard”. The field indeed has made rapid strides in this direction, however, since each ncRNA tends to occupy its own niche in biology (as each new class seems to be a “one-off”), mechanistic studies that hunt down each phenotypic function are slow to work through the as many as ~17 000–96 000 human IncRNAs known to date (Figure 1). In addition, new classes of short ncRNAs, defined as under 200 nucleotides in length, beyond miRNAs are also still being discovered. Perhaps skeptics need to cut the RNA field some slack given that we still do not know the function of a large fraction of the under 20 000 protein-coding genes after over 100 years of study, taking the seminal work on enzymes by Leonor Michaelis and Maud Menten in 1913 as the starting point. In fact, the ~20 years since the human genome project first succeeded in revealing all these ncRNAs seems like a comparably short time to have revealed as much as the field has (Figure 1).
It is also instructive to recall the origin of the term “junk” DNA, coined over 50 years ago by Susumu Ohno based on theoretical considerations of mutational load and the large number of repetitive sequences derived from transposable elements that were assumed to be non-functional. At the time, it was not yet clear that the phenotypic consequences of mutations in protein-coding housekeeping genes can well be catastrophic, while simultaneously the quantitative trait variations affected by mutations in the regulatory architecture that controls these genes during development and beyond can be beneficial by driving the evolution and diversification of complex organisms.

Does this mean that a simple agreement to be patient until elemental function can be expanded into phenotypic function is all that is needed to resolve what remains of the ENCODE controversy?

**INTO THE WEEDS: DEEP-ROOTED ARGUMENTS SUSTAIN SKEPTICISM TOWARD ncRNA FUNCTION**

To better understand and address the strong reaction to the ENCODE project’s conclusions we have to consider several deep-rooted, discipline-informed arguments of the skeptics. In the sections that follow, each major argument is enumerated and thoroughly discussed.

**There is no phenotype!**

The results of the project showed that an elemental function of many ncRNAs is the binding of specific transcription factors. As the example of PRC2 indicates, such binding of specific protein factors can profoundly affect transcription levels, although finding a specific phenotype for such an ncRNA will take time. In fact, genetically deleting an ncRNA may not immediately lead to an observable phenotype. That fact, however, is typical of complex, critical biological functions that often involve subtle functional effects (such as that of individual miRNAs), cell-type- and stress-specific functions, or redundant pathway components to ensure a robust biology. Furthermore, a highly transcribed ncRNA can, simply by binding a specific protein factor, deplete it from its pathways to affect a broader function; in the context of miRNAs, this is called the competing endogenous RNA (ceRNA) hypothesis. Another potential example is found in metastasis-associated lung adenocarcinoma transcript 1 (MALAT1), which is one of the most abundant IncRNAs, highly conserved among mammals, and thought to regulate alternative pre-mRNA splicing and gene expression at both the transcriptional and post-transcriptional levels in a context-dependent manner. While Malat1 knockout mice develop and grow normally, and do not show alterations in alternative splicing, this IncRNA has been associated with the progression of many diseases, in particular cancers, supporting a profound function in the RNA metabolism of the cell, likely mediated by its many associated protein partners. Other IncRNA examples include Neat1, which is involved in placental biology and cognitive function and BC1, implicated in mammalian behavior and synaptic plasticity, learning and memory. In fact, recent high-throughput knockout or knockdown studies of IncRNAs have started to reveal a wide range of cellular and developmental phenotypes, with one study identifying almost 500 IncRNAs required for normal cellular proliferation, of which 89% were expressed in just a single cell type. Clearly, absence of a deletion phenotype in one cell type under optimal conditions is not evidence of the absence of function in all cell types under all conditions.

**Many IncRNAs are sparse!**

It has been noted that fewer than 1000 IncRNAs are present at greater than one copy per cell in the typical human tissue culture cell line, suggesting that the vast majority of ncRNAs is too low in expression to be functional. However, this argument ignores the fact that RNA transcription is highly cell cycle and stress dependent opening the possibility that many IncRNAs may function primarily in single cells at critical junctures. Classic cell culture and tissue analyses, wherein millions of cells are combined to yield enough material for RNA sequencing, average out any outlier cells, and typically do not probe the vast number of stresses that cells may encounter in the human body. Notably, many stress conditions during disease are associated with a significantly changed transcriptome, including many altered ncRNA levels. Similarly, many IncRNAs are transcribed highly, but only transiently, from a plethora of enhancer sequences in the genome to control the time-limited expression of distal genes, a mechanism that guides a cell through its developmental stages. RNA sequencing naturally has not yet been able to probe the transcriptional states of the trillions of cells along their developmental stages during ontogeny of the human body.

A single ncRNA molecule within a crowded cell may only make negligible functional contributions, but of course there are exceptions to this rule, such as genomic DNA molecules. Furthermore, IncRNAs may represent only 1%-4% relative to the ~500 000 mRNA molecules of a single cell, however, other ncRNAs, in particular ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs), outnumber mRNAs by a similar ~40-fold—so are mRNAs non-functional? Finally, IncRNAs often are confined to the nucleus, the root of gene expression, where they only need to bind to one or two specific genomic loci to have a major impact on the expression of nearby genes. Therefore, high expression levels are not needed, in contrast to, for example, a housekeeping mRNA that needs to saturate the cell’s large number of ribosomes for efficient, competitive translation.

**But the C-value paradox!**

Intuitively, one may expect the human genome to be larger than that of onions or amoebae. As it turns out, however, the onion with 16 billion base pairs and the amoeba with 686 billion base pairs have ~5- and ~200-fold larger genomes, respectively, than humans. In fact, genome size varies more than 200 000-fold among eukaryotes. Similarly, even the same genus can show large variation in genome size,
The lack of ncRNA conservation is due to neutral evolution and absence of purifying selection!

Some evolutionary biologists and philosophers have suggested that sequence conservation among genes should be the primary, or perhaps only, criterion to identify functional genetic elements.\(^{[74,75]}\) This line of thinking is based on 50 years of success defining housekeeping and other genes (mostly coding for proteins) based on their sequence conservation. It does not, however, fully acknowledge that evolution does not actually select for sequence conservation. Instead, nature selects for the structure, dynamics and function of a gene, and its transcription and (if protein coding) translation products; as well as for the inertia of the same in pathways in which they are not involved. All that, while residing in the crowded environment of a cell far from equilibrium that is driven primarily by the relative kinetics of all possible interactions.\(^{[24]}\) Given the complexity and time dependence of the cellular environment and its environmental exposures, it is currently impossible to fully understand the emergent properties of life based on simple cause-and-effect reasoning.\(^{[26]}\) Consequently, the estimates of the fraction of the human genome that carries function is still being upward corrected, with the best estimate of confirmed ncRNAs now having surpassed protein-coding genes,\(^{[12]}\) although so far only 10%–40% of these ncRNAs have been shown to have a function in, for example, cell morphology and proliferation, under at least one set of defined conditions.\(^{[50–55]}\) Muddying the waters even more, some ncRNAs encode non-canonical micropeptides less than 100 amino acids in length with diverse and important biological roles,\(^{[76,77]}\) which can result in dual functions in both coding and regulation.\(^{[78]}\) Conversely, protein-coding genes can express functional IncRNAs upon alternative splicing.\(^{[79,80]}\) Finally, the example of the IncRNA component of telomerase—the enzyme that maintains the protected chromosome ends, or telomers, indispensable for genome integrity—illuminates that widely varying RNA sizes and sequences can all be compatible with a structural topology that is conserved functionally from yeast to mammals.\(^{[81]}\)

One example for the scope and limitations of using conservation for defining function are riboswitches, highly structured RNA motifs that are primarily found in bacteria and each bind a specific ligand of cellular importance such as an amino acid, metabolite or elemental ion.\(^{[82]}\) Once the ligand is recognized by the so-called aptamer domain, the interlaced downstream expression platform of the riboswitch changes conformation to module the likelihood for transcription and/or translation (Figure 2A).\(^{[83]}\) Over the past 20 years, Breaker and coworkers have had great success in discovering novel riboswitch classes through their aptamer’s sequence conservation among bacteria of known genomic sequence (Figure 2B), leading so far to over 55 distinct riboswitch classes with at least some bioinformatic, genetic, or biochemical data to validate their functions.\(^{[84]}\) Predictions based on power law projection suggest, however, that up to 28 000 additional, less well-represented riboswitch classes may still be hidden in the genomes of bacterial species that have fewer representatives or are more difficult to culture.\(^{[84]}\) This projection illustrates one limitation of using conservation as a basis for discovering ncRNAs of defined function, namely the limited statistical power to capture less widely distributed sequences than housekeeping genes.

A second limitation of a sequence-based conservation analysis of function is illustrated by recent insights from the functional probing of riboswitches. RNA structure, and hence dynamics and function, is generally established co-translationally, as evident from, for example, bacterial ncRNAs including riboswitches\(^{[85–90]}\) and ribosomal RNAs,\(^{[91,92]}\) as well as the co-transcriptional alternative splicing of eukaryotic pre-mRNAs, responsible for the important, vast diversification of the human proteome across ~200 cell types by excision of...
Riboswitch mechanism, function and conservation. (A) Riboswitches are highly structured RNA motifs embedded in the 5′ untranslated regions of many bacterial mRNAs, where they can enhance or suppress gene expression as they co-transcriptionally fold to bind a small-molecule or elemental ion ligand. One such mechanism involves the modulation of transcription yield by RNA polymerase (RNAP), whereas others more directly alter the likelihood of mRNA translation into protein. (B) The upstream aptamer region binds the ligand, rendering the core segment that forms the binding pocket (yellow box), as well as flanking architectural segments (blue box), highly conserved. By comparison, the downstream expression platform shows much less conservation, most likely because it functionally interacts with many protein effectors that are idiosyncratic to the specific bacterium. Created with BioRender.com.

varying ncRNA introns. In the latter case, it is becoming increasingly clear that splicing regulation involves multiple layers synergistically controlled by the splicing machinery, transcription process, and chromatin structure. In the case of riboswitches, the interactions of the ncRNA with its multiple protein effectors functionally engage essentially all of its nucleotides, sequence-conserved or not, including those responsible for affecting specific distances between other functional elements. Consequently, the expression platform—equally important for the gene regulatory function as the conserved aptamer domain—tends to be far less conserved because it interacts with the idiosyncratic gene expression machinery of the bacterium (Figure 2B). Consequently, taking a riboswitch out of this native environment into a different cell type for synthetic biology purposes has been notoriously challenging. These examples of a holistic functioning of ncRNAs in their species-specific cellular context lay bare the limited power of pure sequence conservation in predicting all functionally relevant nucleotides.

It seems plausible, if not likely, that the genomes of individual species similarly work holistically. In such a model (Figure 3), non-conserved ncRNAs would simply be idiosyncratic imprints of past chromosome or gene duplications, viral infections, transposition events etc., that were accumulated over millennia and at present have been integrated into the overall functioning of the organism. In this context, it is important to recall that the generation time of humans is measured in tens of years, in contrast to the tens of minutes for bacteria, leading to vastly different timescales of evolution, which renders the detection of human genome conservation much more challenging (and the predictive power of such conservation so much weaker). Notably, however, the human body contains ~30 trillion cells, of which ~300 billion are destroyed daily since they have stopped functioning properly. Combined with the observed genomic mosaicism of our body’s cells, enhanced by ncRNAs, these observations imply that the sequenced genome of an individual human being itself is the outcome of relentless internal evolutionary pressures to conform to the needs of all bodily functions. Such considerations are absent from arguments invoking the inability of natural, purifying selection to weed out mildly deleterious mutations among the relatively small size (compared to the human genome) of the human population, which regards only the entire organism as the target of evolution, therefore likely inaccurately suggesting that neutral evolution and drift give rise to the many genomic ncRNAs.

Furthermore, the error frequency of organismal genome replication is generally inversely proportional to their genome length, from humans to bacteria and even viruses, an observation termed Drake’s
Human evolutionary biology, genetics, molecular biology and biochemistry complement each other. (A) The DNA of extant species carries numerous evolutionary imprints of past genetic events that include, for example, a parasitic DNA (pink) transpositioning into the genome and expanding it. While evolutionary biology studies the organismal level, genetics focuses on the genomic DNA tightly packaged within the nucleus of the organism’s cells. (B) Reaching into the nanoscale, molecular biology and biochemistry seek to understand the molecular mechanism of RNAs and their protein partners, where a transposon (pink) may give rise to an RNA transcript that acquires specific functions by forming a complex network of interactions (red dashed box). A functional pathway is characterized by interactions that are kinetically more stable than the alternatives (double-arrows).

One final argument of geneticists is the suggestion that many random pieces of DNA can promote transcription by recruiting transcription factors locally. However, rarely is the transient binding of a single transcription factor sufficient to recruit an RNA polymerase molecule to a transcription site. Rather, a combinatorial cooperation between cis-regulatory sequence elements in the genome, trans-acting transcription factors and signaling molecules, and gene-distal, but cis-acting ncRNA enhancer transcripts is needed to initiate directional transcription events that govern the tissue-specific, spatiotemporally controlled expression dynamics of genes (Figure 3B).

In fact, a majority of lncRNAs appear to arise from genetic loci that act as enhancers that dynamically and adaptably control assembly of the transcription machinery for the spatiotemporal expression of specific genes during the development of multicellular organisms. These enhancer RNAs, or eRNAs, exhibit cell-type specific expression patterns so that they are underrepresented in many RNA sequencing datasets, and remain largely uncatalogued and of poorly defined

Transcription factors bind to random genome sequences!

rule and likely caused by genomes operating near their maximum informational storage capacity and mutation rate. One insight from the COVID-19 epidemic has been the sheer speed with which viruses that produce \( \sim 1 \) trillion viral particles in a human body can evolve. We have to assume that such evolution also occurs at the level of the human body and genome, just at a comparably glacial pace that is difficult to observe. It is therefore plausible to suggest that the human genome is selected equally on a per-nucleotide basis, just as the viral genome, even if we do not yet understand the holistic function(s) of each sequence element.
function[101] Nevertheless, their large number, estimated to be in the hundreds of thousands,[100] contribution to super-enhancers that involve the rapid, transient transcription of eRNAs and their phase separation into eRNA-protein granules, fast evolution under selection for adaptive radiation, and the fact that the vast majority of single nucleotide polymorphisms in the human genome with causal links to complex diseases map to eRNA genes lend them functional relevance.[101–104] Furthermore, the observation that most eukaryotic transcription factor proteins have confusing and enigmatic functions, as well as ill-defined structures, suggests that they have been interpreted using a conceptual framework that misses cooperating eRNAs as a critical link.[40]

In this modern view of eukaryotic gene expression, only those transcription events will occur that are sufficiently robustly proofread by a sequence of kinetically controlled, reversible assembly events that have to enhance each other and outcompete a vast number of possible alternative events.[26] In the resulting holistic model (Figure 3B), the significant number of defined transcripts detected by ENCODE then become a signature of select cellular processes that are allowed to proceed among a much larger number of possible transcripts. While we still do not understand the phenotypic functions of a majority of these primarily non-protein coding RNAs, we have to assume that the likelihood is high for eventually finding many functions that evolution has preserved across the many generations of individuals. Collectively, all these organisms and their cells were exposed to a vast array of rapidly changing environmental conditions, which imprinted on their genome and were inherited by the following generations.

CONCLUSION: HOW TO RECONCILE SCIENTIFIC FIELDS

Science thrives from integrating diverse viewpoints—the more diverse the team, the better the science.[107] Previous attempts at reconciling the divergent assessments about the functional significance of the large number of ncRNAs transcribed from most of the human genome by pointing out that the scientific approaches of geneticists, evolutionary biologists and molecular biologists/biochemists provide complementary information[42] was met with further skepticism.[74] Perhaps a first step toward reconciliation, now that ncRNAs appear to increasingly leave the junkyard[35] would be to substitute the needlessly categorical and derogative word RNA (or DNA) “junk” for the more agnostic and neutral term “ncRNA of unknown phenotypic function”; or “ncRNA_upf.” After all, everyone seems to agree that the controversy mostly stems from divergent definitions of the term “function”[42,74] which each scientific field necessarily defines based on its own need for understanding the molecular and mechanistic details of a system (Figure 3). In addition, “of unknown phenotypic function” honors the null hypothesis that no function manifesting in a phenotype is currently known, but may still be discovered. It also allows for the possibility that, in the end, some transcribed ncRNAs may never be assigned a bona fide function.

Most bioscientists will also agree that we need to continue advancing from simply cataloging non-coding regions of the human genome toward characterizing ncRNA functions, both elementally and phenotypically, an endeavor of great challenge that requires everyone’s input. Solving the enigma of human gene expression, so intricately linked to the regulatory roles of ncRNAs, holds the key to devising personalized medicines to treat most, if not all, human diseases, rendering the stakes high, and unresolved disputes counterproductive.[108] The fact that newly ascendant RNA therapeutics that directly interface with cellular RNAs seem to finally show us a path to success in this challenge[109] only makes the need for deciphering ncRNA function more urgent. Succeeding in this goal would finally fulfill the promise of the human genome project after it revealed so much non-protein coding sequence (Figure 1). As a side effect, it may make updating Wikipedia and encyclopedia entries less controversial.

ACKNOWLEDGMENTS

N.G.W. appreciates support from NIH R35 grant GM131922, as well as thanks two anonymous reviewers for their very valuable feedback and input.

CONFLICT OF INTEREST STATEMENT

The author declares no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Nil G. Walter  https://orcid.org/0000-0002-7301-1275

REFERENCES


How to cite this article: Walter, N. G. (2024). Are non-protein coding RNAs junk or treasure?. BioEssays, e2300201. https://doi.org/10.1002/bies.202300201